PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
H04N 7/18

A1

(11) International Publication Number: WO 94/17636

(43) International Publication Date: 4 August 1994 (04.08.94)

(21) International Application Number: PCT/US94/00866

(22) International Filing Date: 21 January 1994 (21.01.94)

(30) Priority Data:
08/011,067 29 January 1993 (29.01.93) US

(71) Applicant: BELL COMMUNICATIONS RESEARCH, INC. [US/US]; 290 West Mount Pleasant Avenue, Livingston, NJ 07039-2729 (US).

(72) Inventor: BIANCHI, Michael, Haysom; 190 Loantaka Way, Madison, NJ 07940 (US).

(74) Agents: YEADON, Loria, B. et al.; International Coordinator, Room 2E-304, 290 West Mount Pleasant Avenue, Livingston, NJ 07039-2729 (US).

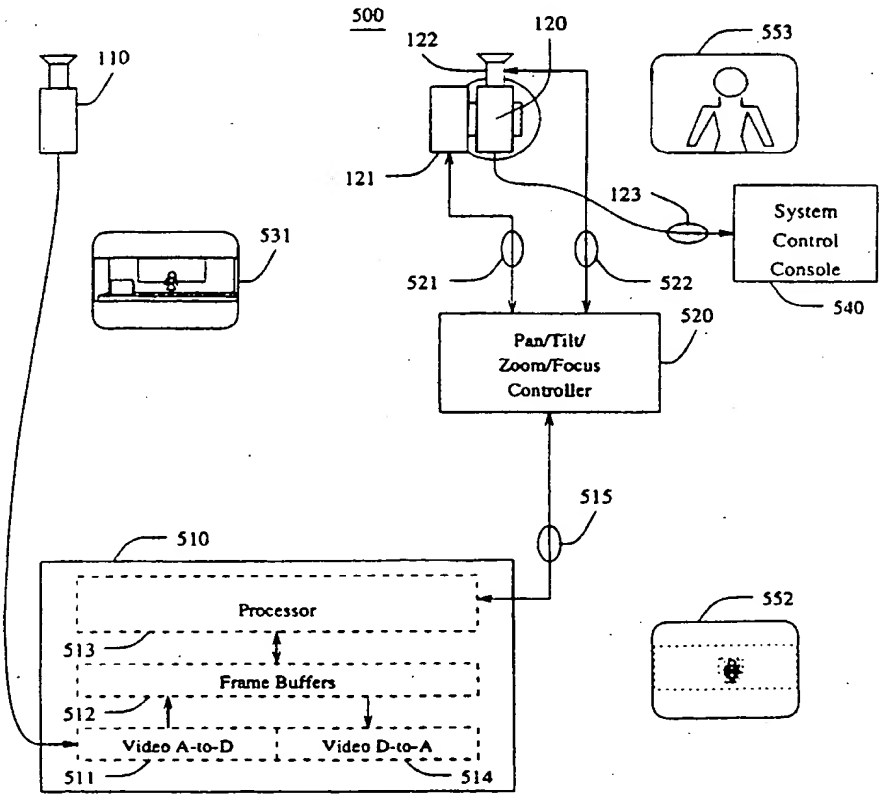
(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: AUTOMATIC TRACKING CAMERA CONTROL SYSTEM

(57) Abstract

Methodology and circuitry for automatically effecting electronic camera movement to track and display the location of a moving object, such as a person presenting a talk to an audience. A fixed spotting camera (110) is used to capture a field of view, and a moving tracking camera (120) with pan/tilt/zoom/focus functions is driven (controller 520) to the present location of the moving object. Information for driving the tracking camera is obtained with reference to the pixel difference between a current image (300) and a previous image (200) within the field of view. A tracking algorithm computes the information necessary to drive the tracking camera from these pixel differences as well as data relative to the field of view of the spotting camera and the present tracking camera position.



BNSDOCID: <WO___9417636A1_I_>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

BNSDOCID: <WO___9417636A1_I_>

AUTOMATIC TRACKING CAMERA CONTROL SYSTEM

Field of the Invention

This invention relates generally to camera control systems and, more specifically, to a two-camera system which automatically tracks movement of an object in the image space under control of a tracking algorithm.

Background of the Invention

Audio/visual presentations given in a corporate setting are seldom attended by as many people as would like to see and hear such presentations. It is often inconvenient to travel to where the talk is given and one must be free at the appointed hour. Televising talks over a video network and/or making the video tape available for later viewing are often viable alternatives for a corporation having many geographical-dispersed work locations.

Installation of commercial-grade television equipment in a large meeting room or auditorium can transform such a location into a simple and cost effective television studio. If the equipment is not overly elaborate nor difficult to operate, then only one person can do the work usually assigned to two or more trained personnel. As a result, it is economical and very convenient to record and telecast presentations made in that room so they can be seen and heard at other locations, and even at different times if desired.

One weakness of the one-operator system is that a person who walks around during their presentation can present a significant work load to the system operator who must keep up with the movement of the person. This extra work load becomes a distraction from the system operator's principal task of presenting the most appropriate image to the remote audience or to the recording medium.

The prior art is devoid of teachings and suggestions for a video system wherein a camera arrangement can track a presenter who paces and/or gesticulates, and thereby provides to the system operator another image which may be appropriately selected for immediate display or recording for later replay.

Summary of the Invention

Instead of requiring the operator to follow the presenter by physically controlling the movement of the camera so as to make available an image at the system control console which may then be displayed to the audience or captured on a recording medium, the technique in accordance with the present invention effects camera movement automatically and the operator merely selects the display image as one of

BNSDOCID: <WO___9417636A1_I_>

many that may be chosen for viewing.

Broadly, in accordance with the methodology of the present invention, a moving object is tracked within the field of view of an electronic camera system, which is illustratively composed of a fixed spotting camera and a movable tracking camera. In the methodology, images are sequentially generated by the electronic camera system and then stored as pixel representations during each of the scan intervals of the camera system. A sequence of pixel differences between pairs of images is then evaluated to produce pixel information indicative of movement of the object. From this pixel information, a set of camera frames is then computed wherein the camera frames are positioned to capture localized movement of the object within the field of view of the electronic camera system. Finally, the each camera frame is imaged with the electronic camera system.

In one illustrative embodiment, the method includes, initially, a step of generating and storing an image, called the previous image, as captured by the camera system during a scan interval. Then, another image, called the current image, is captured during a subsequent scan interval. The difference between the current image and the previous image is evaluated on a per pixel basis to obtain pixel differences. The pixel differences are used to locate a bounding box which outlines or borders the region within the image of the spotting camera indicative of the movement by the object between the current and previous images. The dimensionality of the bounding box is then utilized to generate a tracking frame positioned on the moving object with the appropriate aspect ratio of the camera system. The actual camera frame displayed by the camera system is obtained with reference to the tracking frame, such as by enlarging the area of the tracking frame. Then the current image is generally stored as a previous image, and the cycle repeats, commencing with the capturing of a new current image.

The organization and operation of the invention will be better understood from a consideration of the detailed description of the illustrative embodiments thereof, which follow, when taken in conjunction with the accompanying drawing.

Brief Description of the Drawing

FIG. 1 depicts a typical environment in which the system in accordance with present invention may operate;

FIGS. 2-4 show, respectively, a previous image of information, a current image of information wherein there is movement by the person relative to the previous image, and a depiction of the difference between the previous and current images;

FIG. 5 depicts an illustrative system architecture in accordance with the present invention;

BNSDOCID: <WO___9417636A1_I_>

FIG. 6 depicts a typical spotting image detected by the spotting camera;

FIG. 7 depicts the typical locations of a Threshold Box, a Maximum Search Box, a Current Search Box, and a Block Box within the spotting image;

FIG. 8 depicts a flow diagram of the tracking algorithm in accordance with the present invention;

FIG. 9 depicts the occurrence of a Bounding Box appearing totally within the Current Search Box and outside any Block Box;

FIG. 10 depicts the occurrence of adjacent Sub-Bounding Boxes, one within a Block Box and the other within the Current Search Box but outside any Block Box;

FIG. 11 depicts the occurrence of spaced-apart Sub-Bounding Boxes one within a Block Box and the other within the Current Search Box but outside any Block Box;

FIG. 12 depicts the arrangement of the Current Search Box, the Tracking Frame, and the Camera Frame; and

FIGS. 13 and 14 depict the direction of movement for the edges of the Current Search Box relative to the Camera Frame and Maximum Search Box for the cases of a high Confidence Measure and low Confidence Measure, respectively, as to the location of the person.

Detailed Description

The basic characteristics of the operating environment where the automatic camera control system is used are as follows:

(1) the object being tracked, such as a person, is the only object likely to be moving in the range of potential scenes, except for pre-determined areas which will have motion but where such motion in isolation is not of interest;

(2) only one object is to be tracked, or multiple objects which generally are in motion simultaneously are to be tracked; and

(3) failure to stay "on target" is not a serious flaw, provided it does not happen too often nor persist too long.

These characteristics then impose some limitations on the operating environment, expressed as:

(i) The background of the area of potential scenes must be highly static; for example, in the illustrative case, no drapes or plants constantly swaying, no windows or doorways that can see outside traffic.

(ii) Objects which may move but are not of interest must be well separated from the area where the tracked object moves from the camera's point-of-view. For instance, the

BNSDOCID: <WO__9417636A1_I_>

members of the audience must be well separated from the area where a person giving the talk stands and moves.

(iii) Areas of potential scenes which may exhibit motion and where the tracked object may also move can be identified and marked with a Block Box. As an example, a person giving the talk may walk in front of a projection screen where slides are shown and changed.

Placement of the Camera System in a Room

The diagram of FIG. 1 is an overhead view of room 100, typical of an environment where tracking camera system 109 in accordance with the present invention is used. Person 101 – the object or target of interest – is standing in the center of stage platform 102, in front of projection screen 103, and is facing audience area 105. Two cameras compose electronic camera tracking system 109, namely spotting camera 110 and tracking camera 120. Typically, camera system 109 is suspended from the ceiling of room 100.

Spotting camera 110 does not move and looks at the entire front of the room, as shown by the dotted lines emanating from camera 110 forming a solid viewing angle 111. Camera 110 does not pan, tilt, zoom or change focus.

Tracking camera 120 is remotely controllable and can be made to pan, tilt, zoom and focus by commands issued by a computer (not shown). These commands cause the image within the solid viewing angle 121, called the "tracking image," of the tracking camera to change and thereby tend to keep the moving person 101 in view. The focus function of the tracking camera may include an automatic focus capability.

The image captured within angle 111, called the "spotting image," determines the coverage of the system and hence the range of possible views which may be seen by tracking camera 120. If person 101 leaves the spotting image the tracking camera will not be able to follow him or her outside of angle 111. As the system tends to keep the tracking image centered on the target, it is possible for part of the tracking image to view portions of the room which are outside of the spotting image when the target is near an edge of the spotting image.

Finding the Moving Target

The requirements that only person 101 moves and that everything else in the view of spotting camera 110 be stationary, combined with the sequential-image nature of television, can be effectively utilized to develop a camera control algorithm to "lock onto" person 101.

A television moving picture is a series of static images, much like motion

picture film. Each television image is made up of lines of pixels where each pixel may be in any of a range of brightness values and color values which may be represented by numbers. By way of a hueristic discussion (this is an idealized description of the process – the actual algorithm is discussed in detail below), a non-moving television camera is considered. If one television image is captured and a second image is captured shortly thereafter, and then the image information of the second is subtracted from the first, all the pixels that represent objects which did not move will be the same and their difference will be zero. Those pixels differences whose brightness or color values changed between the first and second image will have non-zero pixel differences. These pixel differences indicate both where the person was and the present location of the (moving) person 101. All pixels whose brightness and color values did not change from the first to the second image represent objects which are not apparently moving and will have pixel differences of zero.

The sequence of images of FIGS. 2-4 depicts the effect. The pictorial information illustrated by spotting image 200 of FIG. 2 represents the first or "previous" spotting image, and the pictorial information illustrated by image 300 of FIG. 3 represents the second or "current" spotting image. "Difference" image 400 of FIG. 4 depicts the absolute value of the pixel differences and, since person 101 was the only object that moved, the double image 405 represents where person 101 was and is. All other objects in the spotting image did not move and the difference image does not represent them.

In practice, the double image 405 itself is not used directly. Instead, as the lines of pixels are processed and corresponding pixels are subtracted, the x and y coordinates of the highest, lowest, left-most, and right-most pixels which were non-zero are marked. These coordinates define the "Bounding Box" 410. The Bounding Box 410 represents the differences of the double image 405 and is used by the tracking algorithm to determine how much panning, tilting, and zooming of tracking camera 120 is appropriate to drive it to tracking image position 420 where it acquires a close-up picture of person 101.

Thus, based on the pixel differences of the previous spotting image 200 and the current spotting image 300 taken by spotting camera 110, tracking camera 120 captures the tracking image of person 101, as discussed in more detail below.

The mapping between the location of the Bounding Box within the difference image 400 (and hence within the current spotting image 200) and the commands sent to Pan/Tilt/Zoom/Focus subsystem (presented below) of tracking camera 120 is based on knowing the settings which aim the tracking camera 120 to predetermined locations within the viewing angle 111 of spotting camera 110. Typically

these predetermined locations may be the four corners of the spotting image 200 of spotting camera 110. These settings are determined during alignment procedures accomplished during system installation. Interpolation allows for the pointing of tracking camera 120 with some accuracy at person 101.

System Architecture

System architecture 500 of an illustrative embodiment in accordance with the present invention is depicted in FIG. 5, and the operation of system 500 is succinctly described as follows:

(1) Video images (an exemplary image is shown by depiction 531) from spotting camera 110 are sequentially captured by Video Analog-to-Digital (A-to-D) converter 511 in computer 510, and stored in one of the frame buffers 512.

(2) Processor 513 retrieves necessary data from a frame buffer 512, analyzes such data, and computes the action, based on a tracking algorithm discussed in detail below, to be effected by tracking camera 120.

(3) As an optional side-effect, processor 513 may develop an image displaying information related to the tracking algorithm. That image is transferred into Video Digital-to-Analog (D-to-A) converter 514 via one of the frame buffers 512 and displayed on a video monitor (not shown); such image display information is shown by depiction 552.

(4) The actions required of tracking camera 120 are communicated to Pan/Tilt/Zoom/Focus controller 520 as commands over interface bus 515. Controller 520 may sometimes respond with position data in response to the commands, and this position data is returned to processor 513 via interface bus 515.

(5) Controller 520 translates those commands into drive signals to perform the Pan/Tilt control and the Zoom/Focus control; these drive signals are transmitted to tracking camera 120 via leads 521 and 522, respectively. The signal on lead 521 is delivered to Pan/Tilt Head 121 of tracking camera 120, whereas the Zoom/Focus signal on lead 522 is delivered to Zoom/Focus subsystem 122 of tracking camera 120. The Pan/Tilt head 121 may sometimes respond with position data in response to the commands, and this position data is returned to controller 520 via leads 521. The Zoom/Focus subsystem 122 may sometimes respond with position data in response to the commands, and this position data is returned to controller 520 via leads 522.

(6) Pan/Tilt head 121 and Zoom/Focus subsystem 122 respond to these control signals, driving the tracking camera to tracking image position 420. The tracking camera 120 thus acquires a tracking image of the person 101 (depiction 553 is exemplary) which is transmitted via lead 123 for use. For example, lead 123 may go

BNSDOCID: <WO___9417636A1_I_>

to system control console 540 where it might become a camera image selected by the operator for display or taping purposes.

Discussion of the Operating Environment

Before the Tracking Algorithm is described in detail, aspects of the operating environment necessary to understanding the algorithm are discussed.

A typical image taken by spotting camera 110, designated spotting image 600, is depicted in FIG. 6. Of particular note is that members 105 of the audience in region 605 are visible within the spotting image and that person 101 can walk in front of the projection screen 103. Both people in the audience and the projection screen may present motion which should normally not be tracked.

Particular areas of interest are designated within spotting image 600, called "boxes", as illustrated in FIG. 7. Specifically, the following boxes are defined: Threshold Box 705, Maximum Search Box 710, and an optional Block Box 715. Each box is defined to the tracking algorithm after spotting camera 110 is installed in the room and remains constant. Defining the boxes is only required once in most installations. In situations where the system must deal with very different environments, such as where the same room has radically different setups and uses, it will be necessary to define several sets of boxes and select the appropriate set for each use. Some uses may require no Block Box 715; others may require more than one Block Box 715.

Threshold Box 705 is placed so that it covers an area of the image which is unlikely to see motion, such as the ceiling of the room. The area within Threshold Box 705 should be illuminated to approximately the same extent as the area within the Maximum Search Box 710.

Maximum Search Box 710 defines the maximum area within the spotting image 600 where control system 109 will attempt to discern motion. Things and people which are likely to move and are not to be tracked, such as members of the audience, should not be within the Maximum Search Box. In FIG. 7, Maximum Search Box 710 is sized from about knee-height of person 101 to as high as the tallest person might reach. The bottom of Maximum Search Box 710 is well above the heads of the audience, so that the heads and raised hands of the audience are unlikely to be within Maximum Search Box 710.

Within the Maximum Search Box 710 is the Current Search Box 720. The Current Search Box is the area within which Pixel Differences are computed and changes in size as the algorithm runs, as described below. It is never larger than the Maximum Search Box 710.

Block Box 715 defines an area at least partially within the Maximum Search Box 710 where there is likely to be motion which is usually, but not always,

ignored. For example, when person 101 changes slides by pushing a button on lectern 104, then changes in the images on projection screen 103 are within Maximum Search Box 710 and therefore would be seen as motion. It is preferable if such motion is not misinterpreted as being the person moving. The algorithm ignores motion within Block Box 715, except when other motion within the Current Search Box is adjacent to it, as described below.

In this illustrative embodiment, each of these boxes is rectangular and there is only one of each. In general, the shape of each area could be arbitrary and there could be multiple instances of each of these within spotting image 600.

Details about the roles of these boxes 705, 710, 715, and 720 will be discussed during the elucidation of the Tracking Algorithm.

The Tracking Algorithm

In order to describe the algorithm concisely, the following terms are defined:

Scan Line Length – the number of pixels in an image scan line.

Pixel Value – the luminance (brightness) and color values of each pixel in the spotting image. Typically luminance values range from zero (representing black) through the maximum pixel value (representing white). A typical maximum luminance pixel value is 255. (The illustrative embodiment does not elucidate use of color information; however, the algorithm may use use color information if it is available.)

Pixel Difference – the absolute value of the difference between two Pixel Values at the same location between the Current Image and the Previous Image.

Rectangle – a rectangular area of the spotting image, defined by the xmin, xmax, ymin, ymax coordinates from the video image wherein: 'x' corresponds to the number of the pixels across a scan line with xmin on the left, xmax on the right; and 'y' corresponds to the number of the scan lines with ymin at the top, ymax at the bottom.

Difference Count – the number of pixels within a Rectangle with Pixel Differences above the Current Threshold.

Bounding Box – the minimum sized Rectangle containing all the pixels with Pixel Differences which are above the Current Threshold.

Sub-Bounding Box – A Bounding Box separated from another Bounding Box; such as a Sub-Bounding Box totally within a Block Box and another Sub-Bounding Box totally outside the same Block Box.

Maximum Search Box – the Rectangle defining the maximum area within the Spotting Image wherein motion is sought.

BNSDOCID: <WO__9417636A1_I_>

Current Search Box – a Rectangle, no larger than the Maximum Search Box and no smaller then the Camera Frame, wherein movement by the object is being searched.

Threshold Box – a Rectangle within the Spotting Image wherein motion is most unlikely; used in determining Current Noise Level.

Current Image – the copy of the lastest digitized spotting camera image stored by processor 513.

Previous Images – copies of earlier, retained digitized spotting camera images stored by processor 513.

Current Noise Level – the largest Pixel Difference in the Threshold Box between the Current Image and the most recent Previous Image.

Threshold Bias – a quantity added to the Current Noise Level as measured within the Threshold Box to obtain the Current Threshold.

Current Threshold – the sum of the Current Noise Level and the Threshold Bias.

Tracking Frame – a Rectangle with the aspect ratio of tracking camera image to which an ideal tracking camera would theoretically be driven.

Camera Frame – a Rectangle with the aspect ratio of tracking camera image to which tracking camera 120 is actually driven. Ideally, the algorithm should drive the tracking camera 120 so that it images all of the Camera Frame and no more.

Confidence Measure – a numerical quantity used as an indication of how much the algorithm believes the current Bounding Box represents the current position of the target being followed. Typically a value between 0 and 100.

Minimum Confidence Level – the Confidence Measure at which Previous Images used in the algorithm may be discarded.

Extra Width Percentage – the fixed percentage added to the width and height of the Tracking Frame to create the Camera Frame.

The following steps determine the initial settings of the Tracking Algorithm and are typically done at system installation.

(a) Mount, align and adjust spotting camera 110 so spotting image 700 images at least the entire area where tracking camera 120 should track person 101.

(b) Mount, align and adjust tracking camera 120 and its pan/tilt head 121 so that they can image all areas of spotting image 700 at the maximum telephoto (greatest magnification) zoom. The algorithm performs best when the lenses of spotting camera 110 and tracking camera 120 are as close to each other as possible. Typical installation places the vertical lens axes so they are coplanar and minimum possible displacement between the horizontal lens axes.

(c) Define Threshold Box 705 to the algorithm.

BNSDOCID: <WO___9417636A1_I_>

(d) Define Maximum Search Box 710 to the algorithm.

(e) Define optional Block Boxes 715 to the algorithm.

(f) Define the correspondence between extreme points within spotting image 700, for instance, the four corners, and the pan/tilt commands sent by processor 513 to the Pan/Tilt/Zoom/Focus controller 520 to point the center of tracking camera 120 image at the corners.

(g) Define the correspondence between the Zoom commands sent by the processor 513 to the Pan/Tilt/Zoom/Focus controller 520 to zoom to the extremes of magnification (maximum wide angle and maximum telephoto image) and the corresponding Camera Frames. Typically this alignment is done with the Camera Frames at the center of the spotting image.

(h) Define the system parameters:

(i) Threshold Bias, typically 10% of the maximum Pixel Value.

(ii) Minimum Confidence level, typically 30.

(iii) Extra Width Percentage, typically 100.

The tracking algorithm is comprised of a loop having the steps discussed in detail below with reference to flow diagram 800 in FIG. 8. For each step of the tracking algorithm, the intent of that step is first described and then the processes which accomplish it are discussed.

In the following steps, "Digitizing a spotting image" implies:

(a) Digitize an image from spotting camera 110 and store in digitized format in a frame buffer 512.

(b) Copy this initial digitized image from a frame buffer 512 into the memory of processor 513.

Processing Block 805: **Initialize the Tracking Algorithm; Generate Previous Image.**

Accomplishes the following:

(a) Set the Current Search Box 720 to the values for the Maximum Search box 710.

(b) Digitize a spotting image as the Previous Image.

These steps ensure that the algorithm is initialized with appropriate values and the processor 513 has at least one Previous Image within its memory to use in the remainder of the algorithm.

Processing Block 810: **Generate Current Image.**

Digitize a spotting image as the Current Image.

This step is required in order to have an image within the memory of processor 513 to compare with the Previous Image.

Processing Block 820: **Compute Current Threshold.**

The Pixel Difference is a combination of actual image differences due to objects moving and video noise. The video noise may be contributed several sources, typically including:

(a) noise in the imaging detector within spotting camera 110,

(b) noise in the imaging circuitry within spotting camera 110, and

(c) noise in video Analog-to-Digital circuitry 511.

If not accounted for, video noise could cause Pixel Differences which would appear to the algorithm as object motion. By calculating a Current Threshold which is higher that the contribution of video noise to the Pixel Differences the algorithm can avoid mistakingly identifying video noise as motion.

As part of the setup procedure, a box in the image of spotting camera 110 is picked as the Threshold Box 705. This box is in a portion of the spotting image 700 of camera 110 which is illuminated approximately the same as the Maximum Search Box 710 but which is never expected to see motion. Therefore, any Pixel Differences within the presumedly static region encompassed by the Threshold Box can be assumed to be due to video noise only. This video noise is assumed to be representative of the video noise which occurs within the Maximum Search Box. The Current Threshold is set as a function of this video noise.

For each pixel within the Threshold Box 710, calculate the Pixel Difference between the Current Image and the Previous Image. The largest value found is the Current Noise Level. Set the Current Threshold as the Current Noise Level plus the Threshold Bias.

Processing Block 830: **Determine Pixel Differences Between Current and Previous Images.**

Creating a double image 405 which is a good indication of where a person 101 is now depends on that person having moved recent enough and far enough for the difference between a Previous Image and the Current Image to create a representative Bounding Box 410. A means to strengthen that difference is to keep several Previous Images, by increasing age relative to the Current Image. If a strong difference occurs between the newest Previous Image and the Current Image, that is an indication that there is a lot of recent motion and is sufficient to continue with this step. If that recent difference shows only slight motion, such as when the person is moving slowly or only moving one portion of their body, comparison between the next oldest Previous Image and the Current Image will create a double image 405 which will show the motion that has occurred over a longer period of time and hence may present a stronger difference on which to base the rest of the algorithm.

BNSDOCID: <WO___9417636A1_I_>

When the person moves in an area only within the Current Search Box, away from any Block Boxes 715, the Bounding Box 410 is defined by the Pixel Differences of his or her double image 405 within the Current Search Box 720. FIG. 9 illustrates this effect, that is, Bounding Box 410 is totally within Current Search Box 720 and totally outside any Block Box 715.

But when person 101 walks in front of projection screen 103, as depicted in FIG. 10, the portion of the double image 405 within the Block Box 715 cannot be ignored. When there is a Sub-Bounding Box 1005 within Block Box 715 and a Sub-Bounding Box 1010 outside the same Block Box 715, and they are adjacent to each other, the two are merged creating the composite Bounding Box 410 used for tracking.

If there is a significant distance between the two Sub-Bounding Boxes, as depicted by FIG. 11, such as when changing the slide generates Sub-Bounding Box 1005 in the Block Box in addition to Sub-Bounding Box 1010 in the outside the same Block Box 715, Sub-Bounding Box 1010 becomes the Bounding Box 410 used for tracking.

Within the Current Search Box 720, calculate the Pixel Difference of the Current Image and the corresponding Previous Image. If the Pixel Difference is greater than the Current Threshold, continue with the following steps:

a. If the pixel is within Block Box 715, record the minimum xmin and ymin and the maximum xmax and ymax pixel coordinates and add one to the count of pixels found above the Current Threshold within that Sub-Bounding Box 1005.
b. Otherwise, record the minimum xmin and ymin and the maximum xmax and ymax pixel coordinates and add one to the count of pixels found above the Current Threshold within that Sub-Bounding Box 1010.

Processing Block 840: **Determine Bounding Box.**

If Sub-Bounding Boxes are found within the Current Search Box and the Block Box, it is now necessary to determine if they should be combined. If the Sub-Bounding Box within the Current Search Box is "near" the Sub-Bounding Box within the Block Box, combine both Sub-Bounding Boxes to form a composite Bounding Box. Otherwise, ignore the Bounding Box in the Block Box. Typically, Sub-Bounding Boxes are considered "near each other" if the number of pixels separating the edges of the Sub-Bounding Boxes are within 1% of the Scan Line Length.

Processing Block 850: **Determine the Tracking Frame.**

Typically, the aspect ratio of the tracking camera image is 4-wide to 3-high. Since a person does not match the 4:3 aspect ratio, a compromise is required when capturing his or her image.

Note that in the operating environment being addressed, the head of person 101 is more appropriate to display than his or her feet. A reasonable compromise

is to position the top of the Tracking Frame 1210 proximate to the top of the Bounding Box (which usually corresponds to the top of the person's head) and make the Tracking Frame as wide as the Bounding Box, as depicted in FIG. 12 .

However, because the algorithm is tracking the difference in the person's position and not the outline of the person, the Bounding Box does not uniformly surround the person's image. The algorithm smooths the size and position of Tracking Frame 1210 to help keep the tracking camera 120 from "jumping around". Many algorithms may be readily devised by those with skill in the art, such as averaging, for smoothing the position of the Tracking Frame. The intent of smoothing the Tracking Frame is to both keep it positioned on person 101, appropriately sized, and to respond quickly when the person moves. The smoothing algorithm may also choose to enlarge the Tracking Frame when the person moves around frequently within a relatively confined area so that the person is always seen in the tracking camera image without moving the tracking camera. Later, if the person stops moving so frequently, the algorithm may choose to shrink the Tracking Frame to create a close-up image. Accordingly, the algorithm is then used to:

Compute the smoothed y position of the top of the Bounding Box. Use the smoothed y as the top of Tracking Frame;

Compute the smoothed width of the Bounding Box. Use the smoothed width as the width of the Tracking Frame; and

Compute the smoothed x position of the center of the Bounding Box. Use the smoothed x as the desired horizontal center of the Tracking Frame.

As part of the smoothing, calculate a Confidence Measure of the difference for use later. This Confidence Measure may based on the number of Pixel Differences that are above the Current Threshold within the Bounding Box. The Confidence Measure may take into account the recent history of the Pixel Differences. If there is a sudden, large increase in the number of Pixel Differences it is likely to be due to an event which is not a person moving, such as a change in room lighting. The Confidence Measure discounts such large changes unless they persist.

Likewise, a very small number of Pixel Differences after a persistent recent history of relatively large numbers of Pixel Differences is likely to be less important and is also discounted, unless they persist for a time.

Processing Block 860: **Drive Tracking Camera to the position defined by the Camera Frame.**

Even with the smoothing described in Processing Block 850, too many fine adjustments to tracking camera 120 image may become a distraction to those watching. As depicted by FIG. 12, hysteresis is introduced by selectively adding a fixed

BNSDOCID: <WO___9417636A1_I_>

percentage (the Extra Width Percentage, typically 100%), to the width and height of the Tracking Frame 1210, creating Camera Frame 1205. Pan/Tilt head 121 and Zoom/Focus subsystem 122 are driven to capture Camera Frame 1205 with tracking camera 120.

Most motions of the Tracking Frame which do not take it outside the Camera Frame do not cause the Camera Frame to move. Motions of the Tracking Frame which do take it outside the Camera Frame cause the Camera Frame to move and hence the Pan/Tilt Head 121 will be moved to point the tracking camera 120 appropriately (see below).

When the Tracking Frame changes size so that the Camera Frame is no longer the fixed percentage bigger than the Tracking Frame, the Camera Frame is adjusted in size to make it conform and hence the zoom lens of the Zoom/Focus subsystem 122 will be adjusted to change the magnification of the image captured by the tracking camera 120 appropriately (see below).

Once the pan, tilt and zoom settings are determined, they are sent to the Pan/Tilt/Zoom/Focus controller 520 which in turn causes the tracking camera 120 to respond appropriately.

When tracking camera 120 and person 101 are both moving, the current automatic focus technology often has difficulty getting a correct setting. Automatic focusing of tracking camera 120 has been much more successful when activated only after tracking camera 120 has stopped moving. Thus, if Camera Frame 1205 has been moving and has now stopped and not moved for a short while (typically, a second or two), processor 513 sends an AutoFocus command to the Zoom/Focus subsystem 122 via the Pan/Tilt/Zoom/Focus controller 520.

Processing Block 870: **Select the Current Search Box**.

When the algorithm has a clear indication of where the person is, the assumption is made that this is the person of interest and there is no need to look anywhere else. Under this condition, the algorithm ignores other motions within the Maximum Search Box but distant from the person by shrinking the Current Search Box to closely surround the person as long as it keeps seeing the person moving. If there is no motion, the algorithm enlarges the Current Search Box looking for motion.

If there are Pixel Differences above the Current Threshold, the edges of the Current Search Box are moved toward the Camera Frame in steps (typically, 5% of the Scan Line Length), as depicted in FIG. 13; otherwise, the edges are moved toward the Maximum Search Box in steps, as illustrated in FIG. 14.

The size of the steps with which the Current Search Box moves toward the Camera Frame and toward the Maximum Search Box need not be the same and may be installation dependent.

Processing Block 880: **Decide Whether or not to Retain the Previous Images.**

The algorithm relies on the image-to-image difference providing enough motion to give a good idea of the person's position. The Confidence Measure expresses the algorithm's assessment of how strongly the algorithm believes it knows the person's current position. If the person is not moving much, the difference may not be very representative of a person and the Confidence Measure would be low. One way to increase the likelihood that there will be significant motion between images and a higher Confidence Measure is to wait longer between images, increasing the time interval between the Previous Images and the Current Image.

In the illustrative embodiment of the invention, the Confidence Measure is based on the Difference Count within the Bounding Box relative to the Difference Counts seen in recent cycles of the algorithm. Other embodiments of the Confidence Measure are also possible.

If the Confidence Measure is lower than the Minimum Confidence Measure the decision is then made to retain the Previous Images. If the Previous Image is retained, then the process continues at Processing Block 810. If the Previous Image is not retained, then the process continues at Processing Block 890.

Also, if the Previous Images are retained for a long period of time, the high Confidence Measure may be due to having a Previous Image which contains a person within the Current Search Box and a Current Image which does not contain a person within the Current Search Box. (This could be the result of a person completely leaving the Current Search Box within the time between the capturing of the Previous Images and Current Image.) For this reason, if the Previous Images have not been updated for a long time, typically 10 seconds, the answer is set to "No" and the process continues at Processing Block 890.

Processing Block 890: **Update Previous Images.**

As seen in Processing Block 830, there may be more than one Previous Image. To update the Previous Images, discard the oldest Previous Image and make the next oldest Image the oldest previous image, and similarly until the Current Image is the most recent Previous Image.

It is to be understood that the above-described embodiments are simply illustrative of the principles in accordance with the present invention. Other embodiments may be readily devised by those skilled in the art which may embody the principles in spirit and scope. Thus, it is to be further understood that the circuit arrangements and concomitant methods described herein are not limited to the specific

forms shown by way of illustration, but may assume other embodiments limited only by the scope of the appended claims.

What is claimed is:

1. A method for tracking a moving object within the field of view of an electronic camera system, the method comprising the steps of

sequentially generating and storing images captured by the electronic camera system during camera scan intervals to produce pixel representations of the images,

generating a sequence of pixel differences between two of the images to produce pixel information indicative of movement of the object,

determining a sequence of camera frames from the pixel information, the camera frames being positioned to capture localized movement of the object within the field of view of the electronic camera system, and

sequentially capturing the camera frames with the electronic camera system.

2. A method for tracking a moving object within the field of view of an electronic camera system, the method comprising the steps of

(a) generating and storing as a previous image the image captured by the electronic camera system during a scan interval,

(b) generating and storing as a current image the image captured by the electronic camera system during a subsequent scan interval,

(c) determining pixel differences between the current image and one previous image,

(d) determining a bounding box from the pixel differences, the bounding box being indicative of the movement detected between the current image and the one previous image,

(e) generating a tracking frame with reference to the bounding box, the tracking frame being indicative of the location of the moving object within the field of view of the electronic camera system,

(f) determining a camera frame from the tracking frame and detecting the camera frame with the electronic camera of the electronic camera system,

(g) storing the current image as the last previous image, and

(h) returning to step (b).

3. The method as recited in claim 2 wherein the electronic camera system is composed of a fixed spotting camera and of a movable tracking camera, wherein

step (a) includes the step of generating and storing as a previous image the image captured by the spotting camera during a scan interval,

step (b) includes the step of generating and storing as a current image the image captured by the spotting camera during a subsequent scan interval,

step (e) includes the step of generating a tracking frame with reference to the bounding box, the tracking frame being indicative of the location of the moving object

BNSDOCID: <WO___9417636A1_I_>

within the image of the spotting camera, and

step (f) includes the step of determining a camera frame from the tracking frame and detecting the camera frame with the tracking camera.

4. A method for tracking a moving object within the field of view of an electronic camera system composed of a spotting camera and a tracking camera, the method comprising the steps of

(a) generating and storing as a previous image the image captured by the spotting camera during a scan interval,

(b) generating and storing as a current image the image captured by the spotting camera during a subsequent scan interval,

(c) determining pixel differences between the current image and one previous image,

(d) determining a bounding box from the pixel differences, the bounding box being indicative of the movement detected between the current image and the one previous image,

(e) generating a tracking frame with reference to the bounding box, the tracking frame being indicative of the location of the moving object within the image of the spotting camera,

(f) determining a camera frame from the tracking frame and detecting the camera frame with the tracking camera,

(g) storing the current image as the last previous image, and

(h) returning to step (b).

5. The method as recited in claim 4 further comprising the step, executed after said step (f), of selecting a current search area from the image captured by the spotting camera to locate the moving object.

6. The method as recited in claim 4 further comprising the step, executed after said step (b), of determining a threshold box within the field of view to produce a current threshold value representative of ambient image conditions, said pixel differences then computed with reference to the current threshold value.

7. A method for tracking a moving object within the field of view of an electronic camera system composed of a spotting camera and a tracking camera, the method comprising the steps of

(a) generating and storing as a previous image the image captured by the spotting camera during a scan interval,

(b) generating and storing as a current image the image captured by the spotting camera during a subsequent scan interval,

(c) determining pixel differences between the current image and one previous

image,

(d) if the pixel differences are below a predetermined threshold, returning to step (b); otherwise, continuing with the next step,

(e) determining a bounding box from the pixel differences, the bounding box being indicative of the movement detected between the current image and the one previous image,

(f) generating a tracking frame with reference to the bounding box, the tracking frame being indicative of the location of the moving object within the image of the spotting camera,

(g) determining a camera frame from the tracking frame and detecting the camera frame with the tracking camera,

(h) storing the current image as the last previous image, and

(i) returning to step (b).

8. The method as recited in claim 7 further comprising the step, executed after said step (g), of selecting a current search area from the image captured by the spotting camera to locate the moving object.

9. The method as recited in claim 7 further comprising the step, executed after said step (b), of determining a threshold box within the field of view to produce a current threshold value represenative of ambient image conditions, said pixel differences then computed with reference to the current threshold value.

10. Circuitry for tracking a moving object within the field of view of an electronic camera system composed of a spotting camera and a tracking camera, the circuitry comprising

means for sequentially generating and for storing, as a previous image, the image captured by the spotting camera during a scan interval and, as a current image, the image captured by the spotting camera during a subsequent scan interval,

means, coupled to said means for sequentially generating and for storing, for determining the pixel differences between the current image and one previous image,

means, coupled to said means for determining, for generating a bounding box from the pixel differences, the bounding box being indicative of the movement detected between the current image and the one previous image,

means, coupled to said means for generating a bounding box, for generating a tracking frame from the bounding box, the tracking frame being indicative of the location of the moving object within the image of the spotting camera, and

means, coupled to said means for generating a tracking frame, for determining a camera frame from the tracking frame, and

means, coupled to said means for determining a camera frame, for capturing the

BNSDOCID: <WO___9417636A1_I_>

camera frame by the tracking camera.

11. The circuitry as recited in claim 10 further comprising means, responsive to the spotting camera and coupled to said means for sequentially generating and for storing, for selecting a current search area from the image captured by the spotting camera to locate the moving object.

12. An electronic camera system for tracking the movement of an object within the field of view of the camera system, the system comprising

a spotting camera for capturing video images within the field of view,

a tracking camera having pan, tilt, zoom, and focus functions,

a tracking camera controller for controlling said functions,

a video controller, coupled to said spotting camera, for receiving and for converting said video images to produce a set of digitized images,

a frame buffer, coupled to said video controller, for storing said digitized images, and

a processor, coupled to said frame buffer and said tracking camera controller, for processing said digitized images to generate pixel differences from said digitized images, said pixel differences being representative of localized movement of the object within the field of view of the camera system, said pixel differences being processed by said processor to produce a sequence of camera frames and, correspondingly, processor control signals for controlling said tracking camera controller, said tracking camera controller responsive to said control signals to drive said pan, tilt, zoom, and focus functions so that said tracking camera captures the moving object.

13. The system as recited in claim 12 wherein said processor includes means for storing and executing a tracking algorithm having said digitized images as inputs and said control signals as outputs.

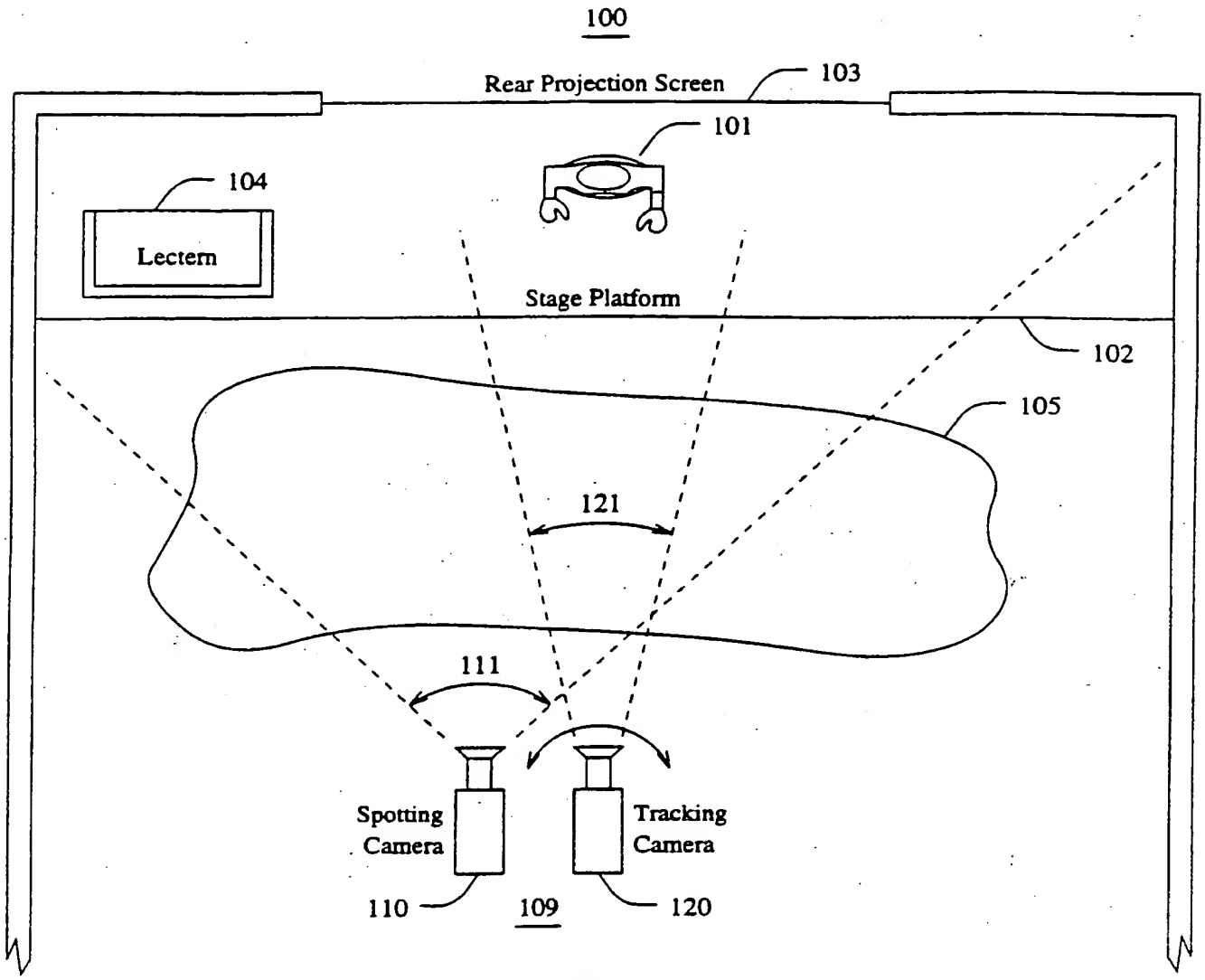


FIG. 1

BNSDOCID: <WO___9417636A1_I_>

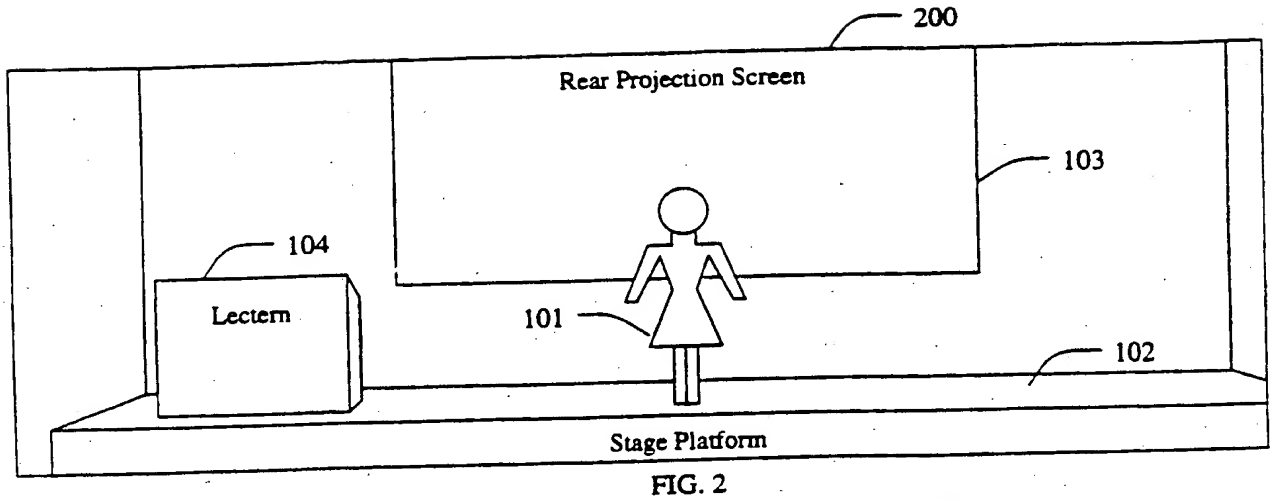


FIG. 2

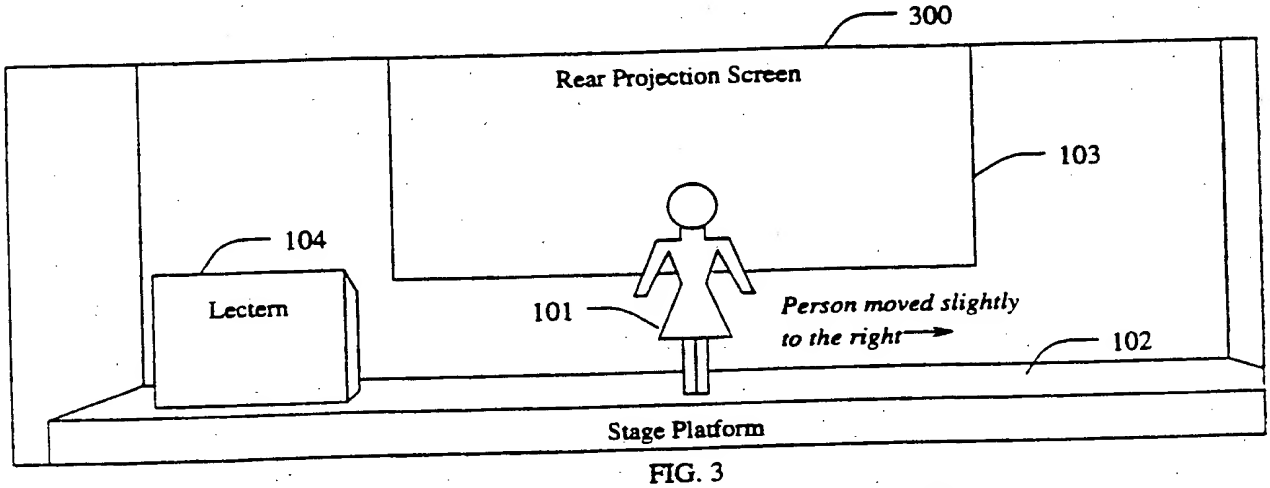


FIG. 3

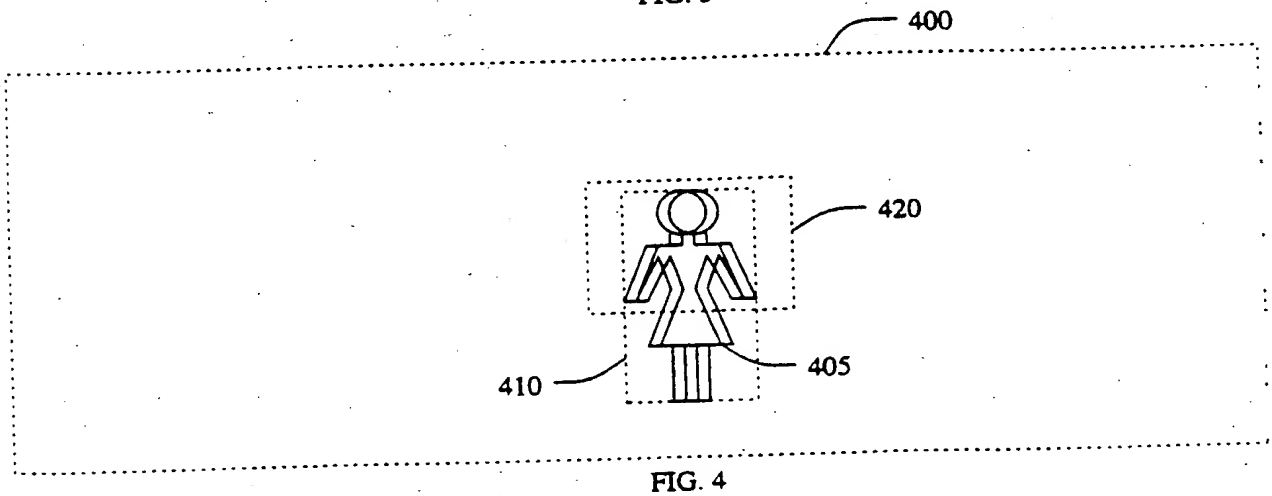


FIG. 4

BNSDOCID: <WO___9417636A1_I_>

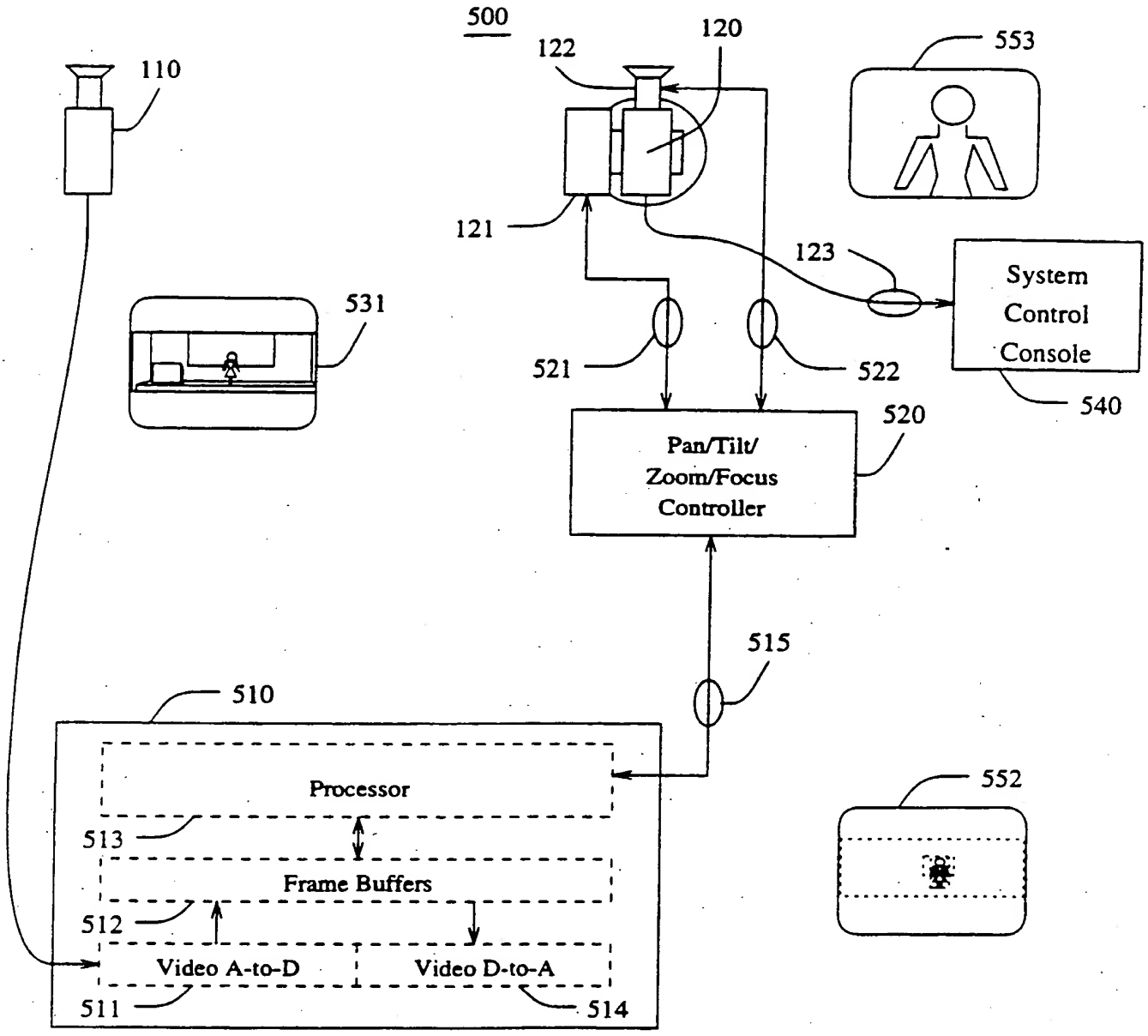


FIG. 5

BNSDOCID: <WO___9417636A1_I_>

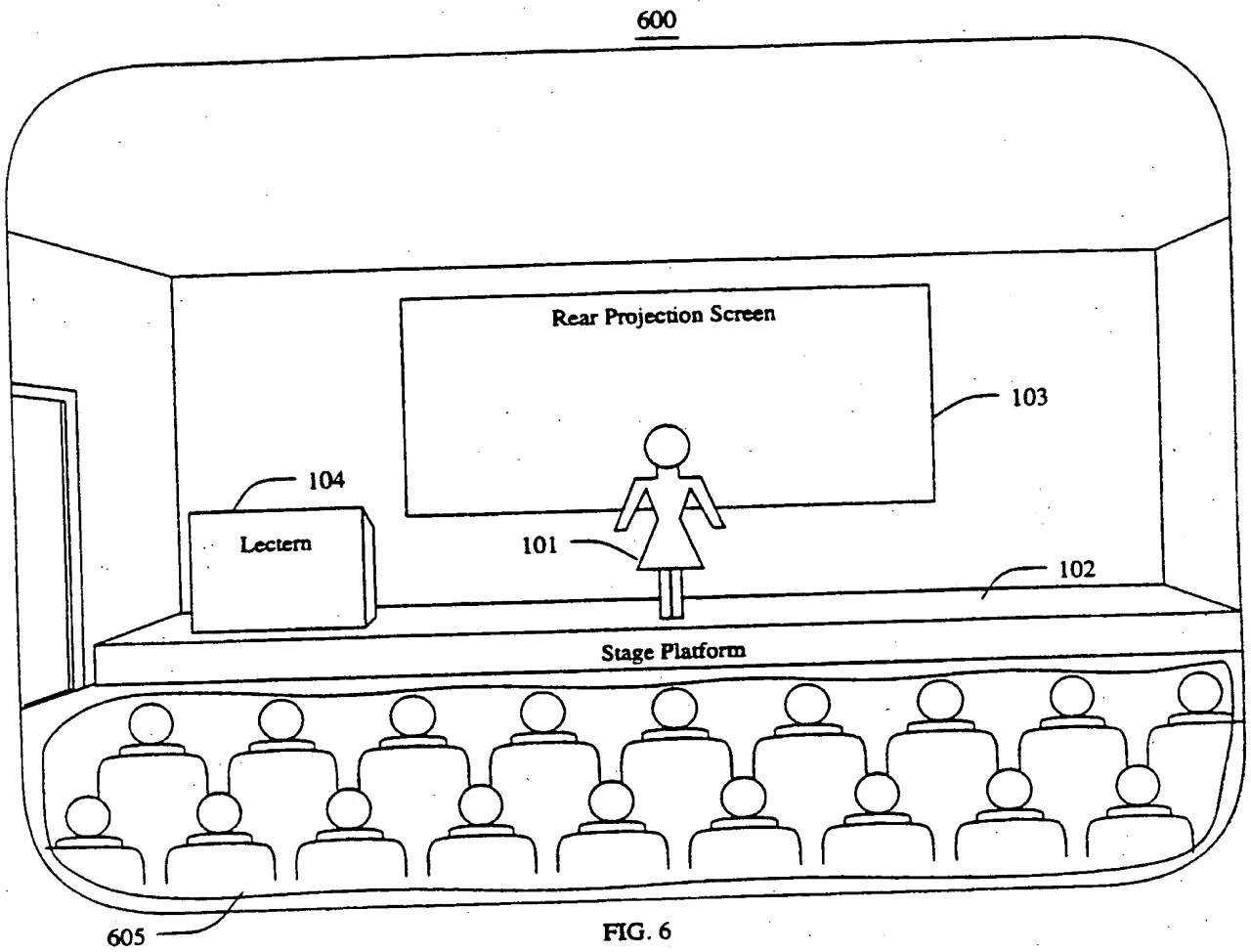


FIG. 6

BNSDOCID <WO___9417636A1_I_>

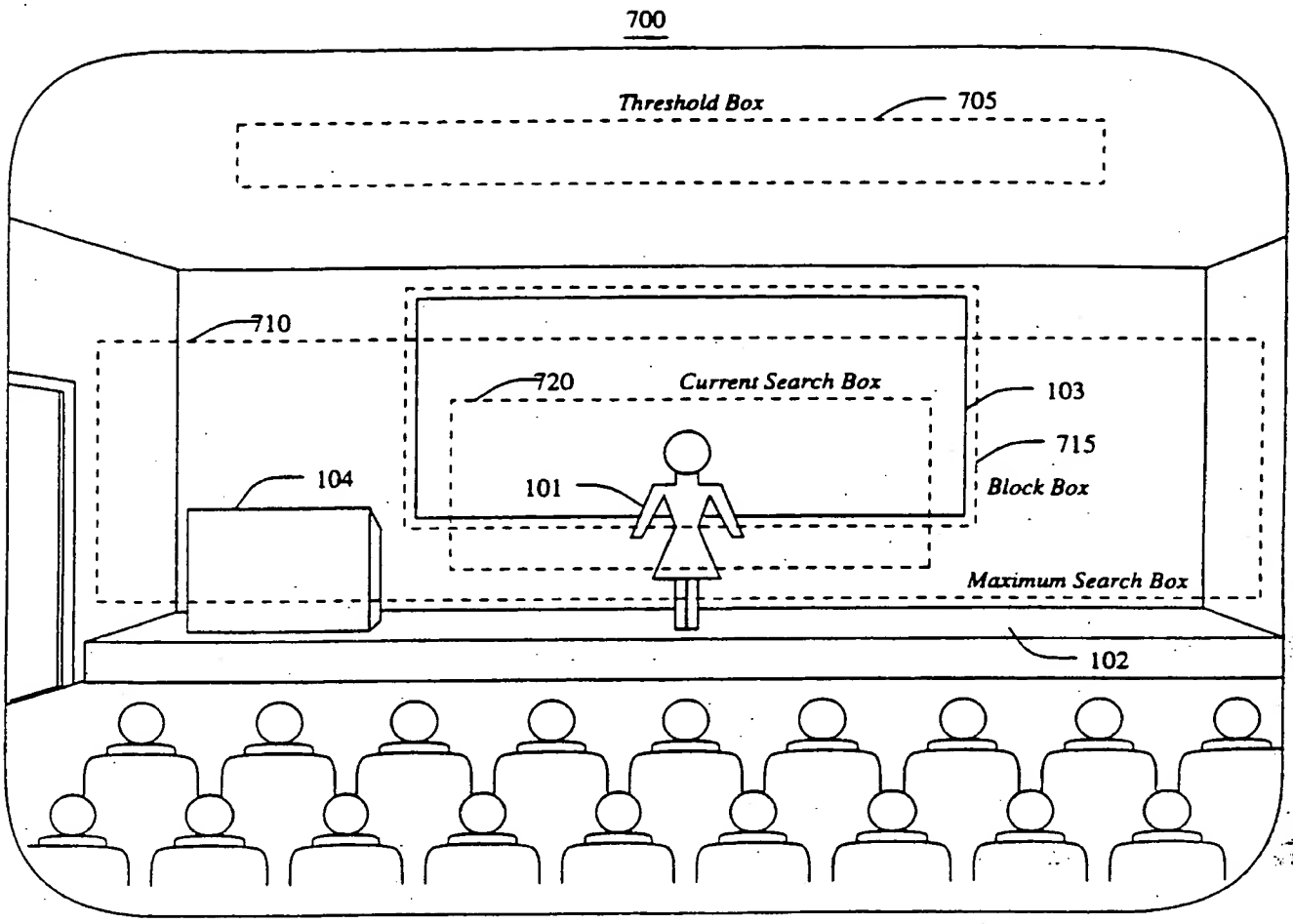


FIG. 7

BNSDOCID: <WO___9417636A1_I_>

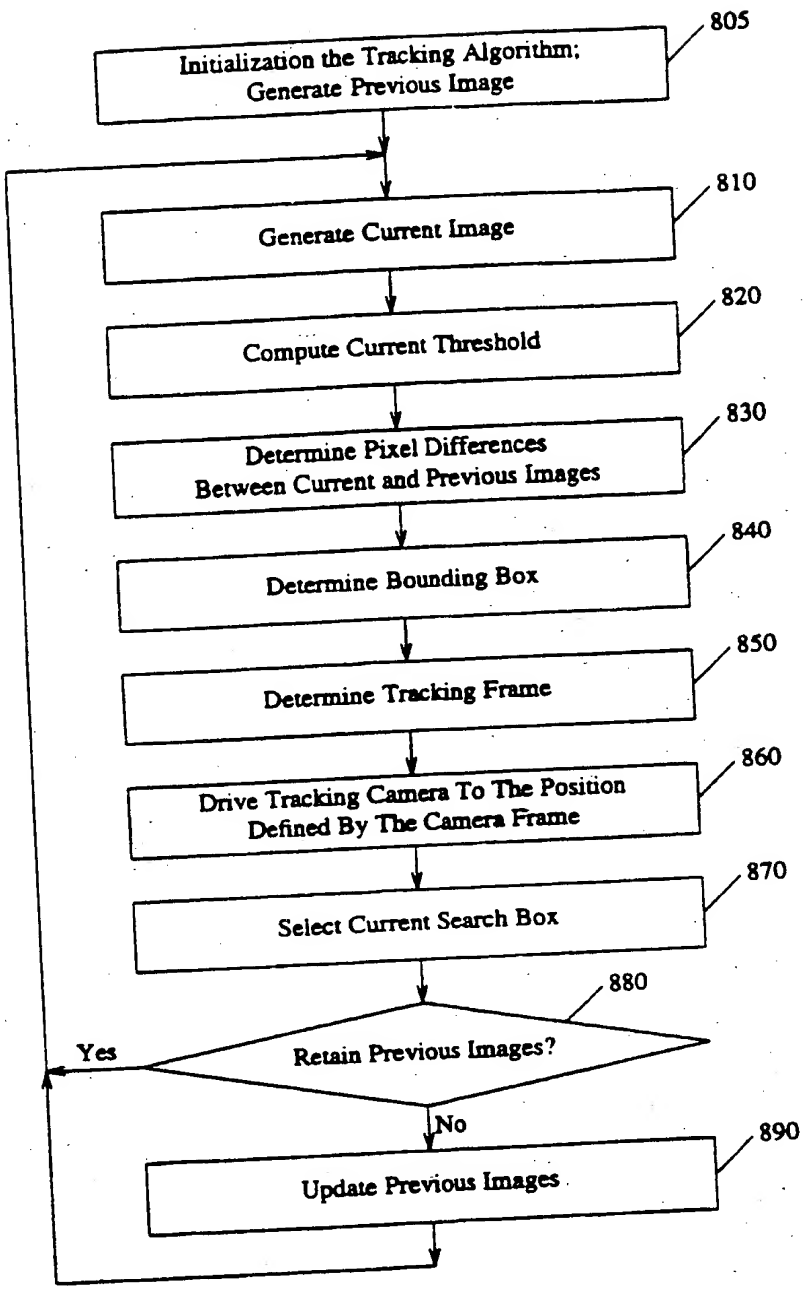


FIG. 8

BNSDOCID <WO___9417636A1_I_>

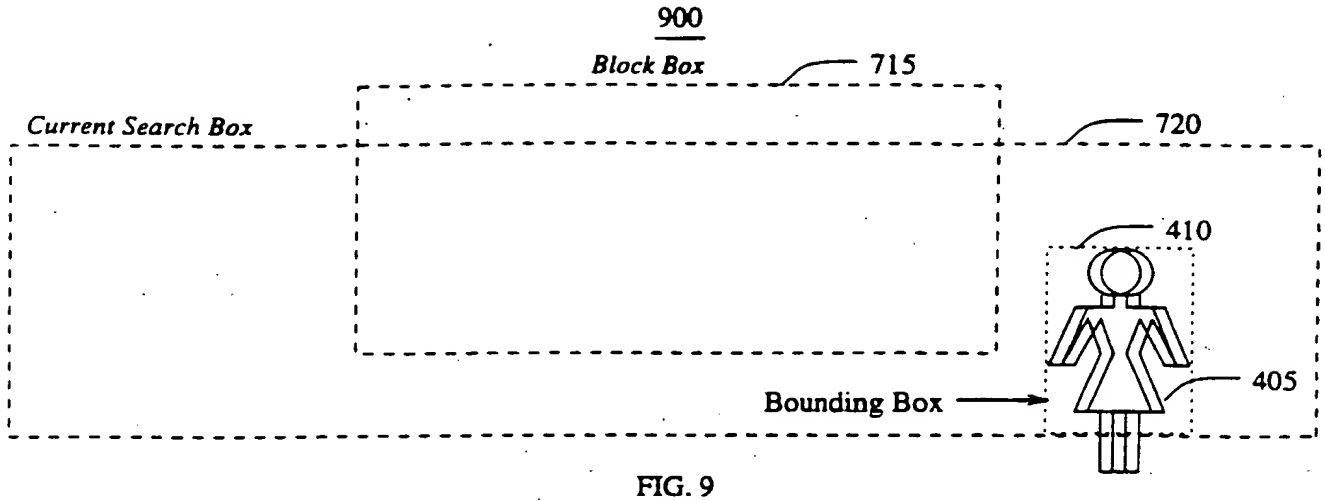


FIG. 9

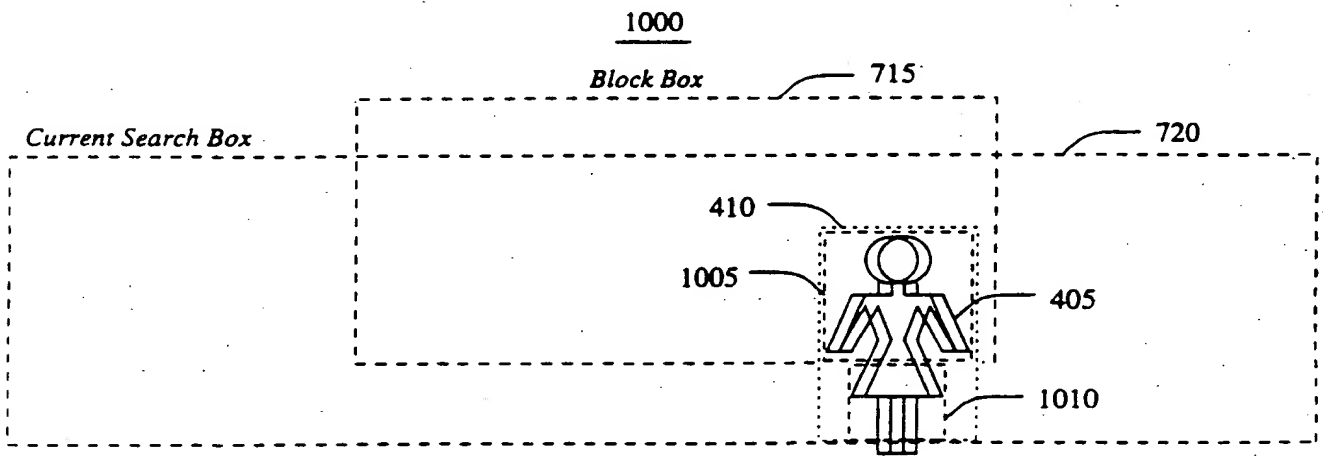


FIG. 10

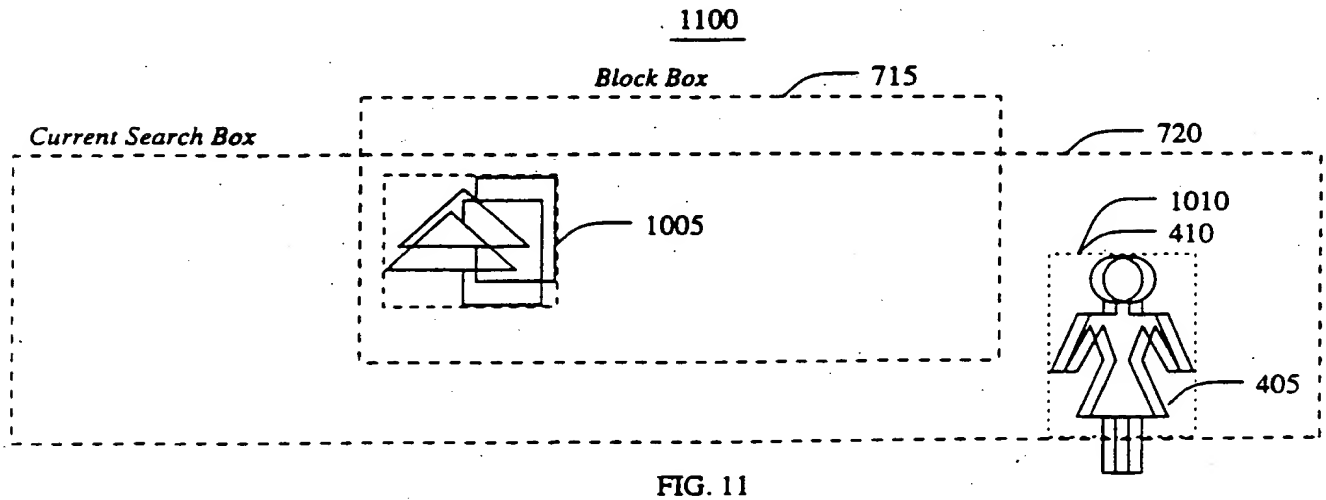


FIG. 11

BNSDOCID: <WO___9417636A1_I_>

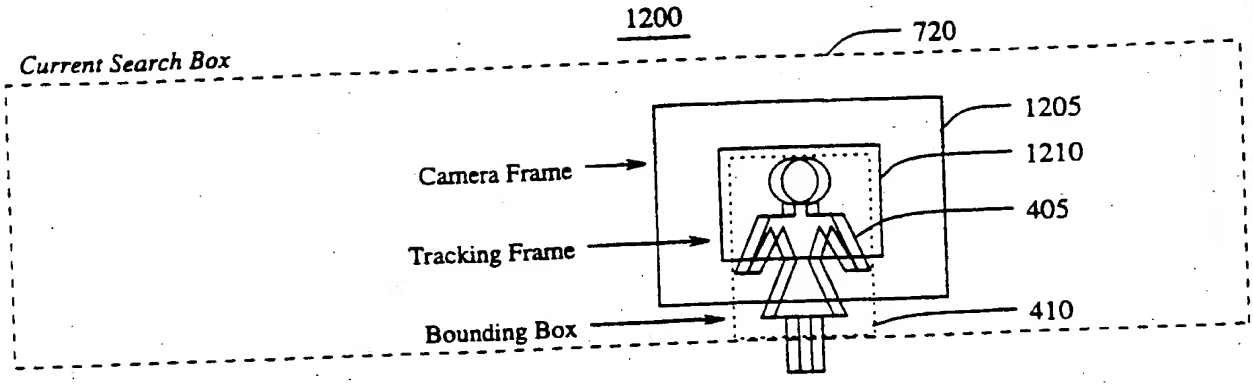


FIG. 12

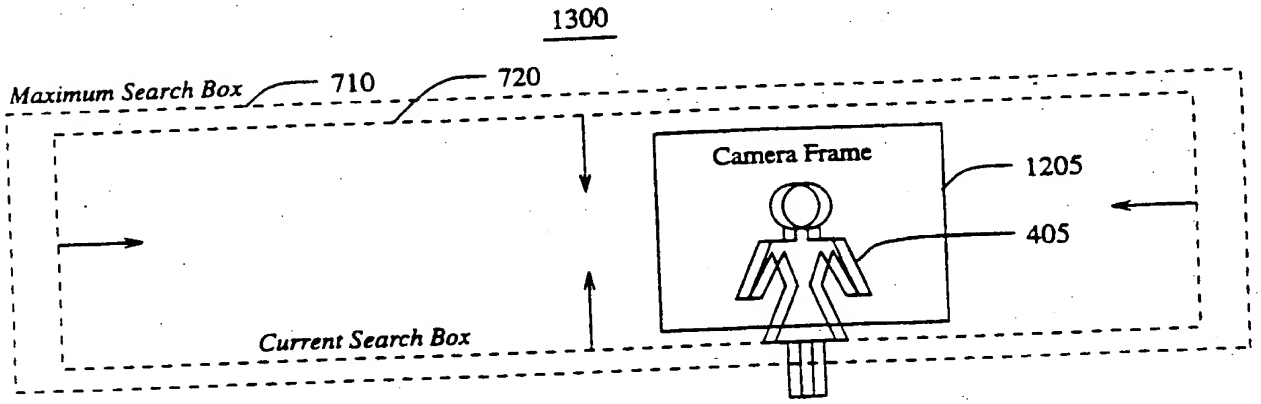


FIG. 13

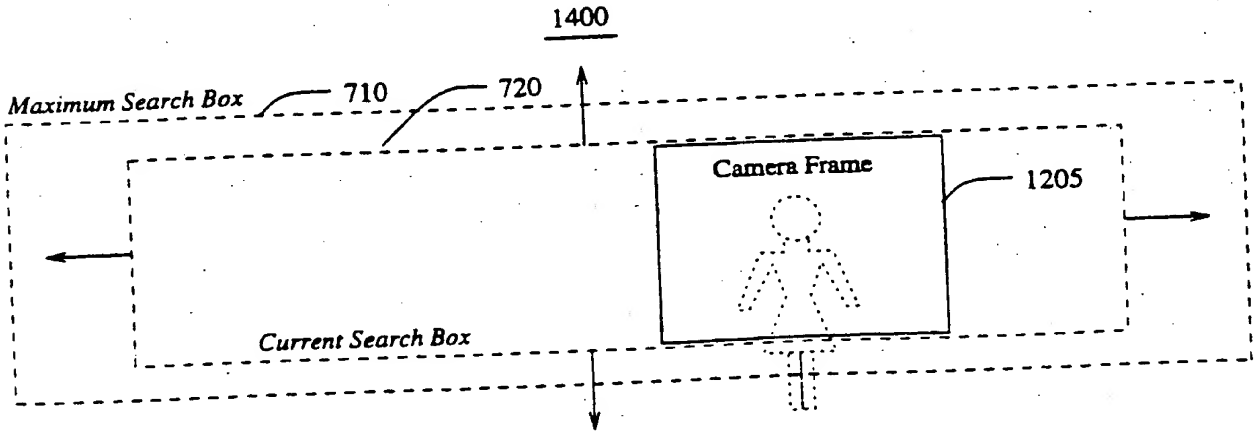


FIG. 14

BNSDOCID <WO___9417636A1_I_>

INTERNATIONAL SEARCH REPORT

[illegible]tional application No.
[illegible]/US94/00866

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :H04N 7/18
US CL : 348/169,170

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 348/143,155,169,170,211; IPC(5) : H04N 7/18

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US, A, 5,243,418 (KUNO ET AL) 07 September 1993, *figure 2, columns 6-7.	1, 2
Y	US, A, 3,652,790 (EBERHARDT) 28 March 1972, *figure 1, column 2.	3-11
Y	US, A, 5,164,827 (PAFF) 17 November 1992, *figures 2, 6.	12, 13
A	US, A, 5,091,781 (AN) 25 February 1992.	1-13
A	US, A, 4,268,863 (LOS) 19 May 1981.	1-13
A	US, A, 4,951,137 (KISOU ET AL) 21 August 1990.	1-13
A	US, A, 4,120,004 (COUTTA) 10 October 1978.	1-13

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 11 APRIL 1994	Date of mailing of the international search report JUN 16 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer RICHARD LEE Telephone No. (703) 308-6612

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO___9417636A1_I_>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/00866

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 5,175,694 (AMATO) 29 December 1992.	1-13
A	US, A, 3,891,798 (MARCUS) 24 June 1975.	1-13
A	US, A, 3,953,669 (SACCOMANI ET AL) 27 April 1976.	1-13
A	US, A, 4,405,940 (WOOLFSON ET AL) 20 September 1983.	1-13

BNSDOCID <WO___9417636A1_I_>